DOCUMENT RESUME

ED 195 357 PS 011 872

AUTHOR Marcus, Robert F.: And Others
TITLE Application of the "Generalizability Coefficient" to the Study of the Stability of Cooperative Behavior of Young Children.
PUB DATE Sep 80
NOTE 17p.: Paper presented at the Annual Convention of the American Psychological Association (88th, Montreal, Quebec, Canada, September 1-5, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavioral Science Research: *Cooperation: Observation: *Play: *Preschool Children: *Prosocial Behavior: Reliability: Sampling: *Statistical Analysis
IDENTIFIERS Cronbach (Lee J): *Generalizability Theory: *Time Sampling

ABSTRACT

Time sampled observations of the cooperative behavior of two samples of 31 preschool children were analyzed for stability (that is, short term reliability of behavior) over a 2-month period using Cronbach's generalizability coefficient. Observations were made during free play periods on nursery school settings. The observation schedule required observers to check the presence of cooperative play, parallel play, and individual play with peers. Only the cooperative play results are analyzed here. The schedule further required that cooperative behavior for each child be time sampled using 10 second intervals for an average of 24 units per day on at least six separate days within a 2-month period. Total observation time per child was 24 minutes. Cooperative behavior in both samples was shown to be highly stable and influenced by the overall amount of cooperative behavior shown but not by sex and age. Results suggest that for research purposes individual difference measures may be based upon naturalistic observations in such settings. Methodological issues related to the reliability of behavior are discussed. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195357

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Application of the "Generalizability Coefficient" to the Study of the Stability of Cooperative Behavior of Young Children

by

Robert F. Marcus, Ph.D.
(147-34-5643)
Charles Johnson, Ph.D.
and
Edward J. Roke, Ph.D.

The University of Maryland
College Park, Maryland 20742

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert F. Marcus
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PS 01872

Abstract

Time sampled observations of the cooperative bahavior of 62 preschool children were analyzed for stability over a two month period using Cronbach's generalizability coefficient. Cooperative behavior in both samples was shown to be highly stable and influenced by the overall amount of cooperative behavior shown but not by sex and age. Methodological issues related to the reliability of behavior are discussed.

Stability of Cooperative Behavior

147-34-5643

Evidence on short term stability of various kinds of behavior measures, sometimes referred to as the reliability of the behavior rather than reliability of observers, has not received as much attention as has long term stability. This is unfortunate since a common strategy for studying social behavior involves observation or other measurement of parent characteristics in order to study their correlation with observational measures of children's social behavior in naturalistic settings. Researchers hope that by following precedent and collecting from a half-hour to an hour's worth of 5 or 10 second interval sampling they will obtain measures which are reliable or generalizable enough measures of individual differences between children to allow conclusions about such relationships.

Generalizability (G) theory (Cronbach et al., 1974) provides a useful framework for studying such stability for several reasons. G theory is less restrictive in assumptions -- occasions or settings are assumed to be randomly sampled from a common universe of settings, but are not assumed to provide parallel measures as is the case with classical reliability theory.

G theory accounts for the existence of interactions of persons and measurement facets such as occasions. Person-occasion interactions in naturalistic observational research tend to be

large enough in many cases to preclude efficient measurement of individual differences (Cairns and Green, 1979). Various theorists have raised substantive issues with regard to the liklihood of the stability of such important behaviors as cooperation within the early childhood years. Kagan (1969) has said that phenotypic continuity is "...likely to be minimal during the opening decade of development and especially during the first five years...(because) the form and eliciting conditions of behavior change so dramatically". Mischel (1969) argues for the stability of cognitive but not personality- social characteristics across situations and over time, unless there is consistency of situational eliciting conditions. One could argue that such homogeneity is indeed present within a nursery school setting in which the people, toys, physical features, time of day, activities, etc. remain essentially the same each day.

G analysis also provides a means of determining whether the size of a particular time sample is adequate for generalizablity to other occasions. Recent comments on methodology in naturalistic observation by Radke-Yarrow and Zahn-Waxler (1979) suggest that in attempting to obtain a sufficient sample of behavor (1) time sampling is probably not suitable for sporadic behaviors, such as tantrums; (2) if certain behaviors occurred only infrequently within even 40 minutes of time per child, that this amount of time would not be sufficient, and (3) if cooperation and helping behavior was sampled within adjacent 5 to 10 minute and 10 to 30 minute periods, the correlations were much lower (and non-significant) than samplings

taken with larger (60 minute) periods up to 2 weeks apart. These guidelines help the investigator decide on the appropriateness of time as a unit of analysis, the adequacy of total time for certain behaviors based upon their total frequency of occurance, and the importance of sampling across broader periods of time. G theory provides a statistical method of detecting when reliability is sufficient and gives a researcher a means of deciding when enough data has been collected during the course of a study to meet some set minimum level of stability. Generalizability analyses from pilot studies or previous research can be used to estimate the size occasion-sample needed to satisfy research requirements.

METHOD

Two samples of 31 children each were observed in two nursery school settings. Sample A consisted of 9 girls and 21 boys with a mean age of 47.7 months. Sample B consisted of 14 girls and 17 boys with a mean age of 59.4 months. Observations were made during free play periods. The observation schedule required observers to check the presence of cooperative play, parallel play, and individual play with peers. Only the cooperative play results are analyzed here. The schedule required that cooperative behavior for each child be time-sampled using 10 second intervals for an average of 24 units per day on at least 6 separate days within a 2-month period (total time per child was 24 minutes). Inter-observer agreement prior to the

study was 90%.

RESULTS

Data for each study consisted of 31x6 subjects by occasions matrices of occasion scores. Each score was the number of intervals in which cooperative behavior was observed on a particular day. A model for subject i's score on occasion J is:

$$X_{ij} = \mu_{..} + S_i + O_j + r_{ij}$$

where $\mu_{..}$ is the grand mean of a subject by occasion universe of scores

S_i is the effect associated with subject i $(\mu_i - \mu_{..})$

O_j is the effect for occasion J $(\mu_j - \mu_{..})$

r_{ij} is the residual $(X_{ij} - \mu_{..} - S_i - O_j)$

and subjects and occasions are assumed to be random samples from a universe of subjects and occasions defined by the researcher.

A two-way ANOVA with one observation per cell was applied to each matrix of scores. Under the random effects model assumed above, the Cornfield-Tukey algorithm was used to estimate variance components of the universe of ratings (Hays, 1973). Tables 1 and 2 present the ANOVA and estimated variance components.

Table 1

Analysis of Variance of Subject by Occasion
Ratings for Studies A and B

	Study A			Study B		
Source	Degrees of Freedom	Sum of Squares	Mean Square	Degrees of Freedom	Sum of Squares	Mean Square
Subjects	30	520.47	17.35	30	300.28	10.01
Occasions	5	27.64	5.53	5	33.96	6.79
Residual	150	721.53	4.81	150	711.20	4.74

Table 2

Estimated Variance Components for Subjects, Occasions and Residual

	Study A		Study B	
Source	Variance Component	Proportional Size	Variance Component	Proportional Size
Subjects	2.090	.302	.878	.154
Occasions	.023	.003	.066	.012
Residual	4.810	.695	4.741	.834

The variance components provide the main basis for the generalizability analysis. They indicate the extent to which variability is produced by various dimensions in the design and sum to a total variance estimate under the model and assumptions above. The ANOVA mean squares provide the basis for their estimation.

Interpretation of Variance Components

The subjects' variance component reflects the extent of individual differences on the measure of cooperation. The subjects' component in study A is more than twice as large as that for study B (it is also about twice as large relative to total variability). This indicates that subjects were more heterogeneous on the cooperation measure in the first study. Since the two samples were obtained from different schools or situations, this difference is not surprising. Although the 15 and 30% of total variance accounted for by the subjects' component may seem small, this relative size can be compared to the typical contribution of the subjects' component on psychological tests. For a 50 item test with an estimated alpha coefficient of .80, the estimated variance component for persons or subjects would account for only about 7% of total response variation. At the individual occasion level, there are too many factors determining the response for subjects influence to be large. At the score level, however, based upon a sample of occasions or days, individual differences among subjects may

still be the largest contributor to score differences.

The occasions variance component indicates the extent to which the overall levels of cooperation varied from occasion to occasion; or the extent to which the occasions on the average tended to elicit different levels of measured cooperativeness. This source contributed relatively little to total variability in the two studies.

The variance component for residuals reflects a) subject by occasion interaction (e.g. the tendency for a generally cooperative subject to respond uncooperatively to particular occasions) and b) the effects of all other unexplained or unknown contributions to score variability (error in the usual ANOVA terminology). The influences of these sources cannot be separately estimated here since there is only one observation per subject by occasion cell in the analysis. If the design had included an additional dimension or facet (such as raters) it would have been possible to obtain a separate estimate of subject by occasion interaction and error would then be confounded with the highest- order interaction in the design (which would be subjects x occasions x raters).

The largest source of variability for the two studies is residual. This is typically the case in such analyses, since the residual source reflects many different contributions. The two resicuals here are similar in both absolute and relative size. The residual source accounts for about 70% of total variability in study A and 83% of total variability in study B.

Estimated Generalizability of Measurement Samples

The estimated generalizability coefficient for a sample of n_o occasions is defined for the design used here as:

$$\hat{\rho}^2 = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \hat{\sigma}_r^2/n_o}$$

where n_o is the number of occasions sampled per subject. The denominator is the expected obtained score variance or variance of subject scores. Note that the influence of the residual is directly dependent upon the number of occasions sampled per subject so that generalizability will increase with larger occasion samples as would be expected. The generalizability coefficients for the 6-occasion measures obtained here (Table 3) indicate that 72% of the score variance in study A reflects individual differences among subjects and 52% for study B. The difference reflects the greater subject individual differences in study A, since the residuals are similar for the two studies. The variance component estimates (Table 2) can be used with the above generalizability coefficient formula to estimate generalizability for measurement samples of other (larger or smaller) sizes by substituting different values for n_o.

Table 3

Estimated Score Components and Generalizability Coefficients for Ratings of Six Occasions Per Subject

Source	Study A	Study B
Subjects	2.090	.878
Residual	.802	.790
Estimated Score Variance	2.892	1.668
Generalizability Coefficient	.72	.53

Table 4

Estimate of Error Variance and the Coefficient of Generalizability for Various Occasion Sample

Number of Occasions	Study A		Study B	
	Error Variance	Generalizability Coefficient	Error Variance	Generalizability Coefficient
2	2.416	.465	2.371	.270
4	1.208	.635	.824	.426
6	.806	.723	.792	.526
8	.604	.777	.593	.597
10	.483	.813	.474	.649
12	.403	.839	.395	.690
14	.345	.859	.339	.721

Table 5

Forced-Order Regression Analysis of Contributions to Measurement Error

Order	Source[a]	Degrees of Freedom	Sum of Squares	Mean Square	F
1	Study	1	.40	.40	1.35
2	Score	1	39.91	39.91	131.03
3a	Age	1	.42	.42	1.40
3b	Residual	58	17.58	.30	
4a	Sex	1	.04	.04	.13
4b	Residual	58	17.96	.31	

[a]Age and Sex were entered at step 3 in separate analyses.
The residual mean-square of .30 was used for testing study and score effects.

Table 4 presents such estimates for various occasions per subject sample sizes. Generally, table 4 indicates that relatively small increases in occasion sample size (e.g. from 6 up to 10) may be profitable here although a point of diminishing returns is reached relatively soon.

Correlates of stability or Measurement Error

The intra-individual standard deviation across the 6 occasions was used as a measure of measurement stability (or individual error). A forced-order regressin analysis (table 5) with stability as the dependent variable was run to study the relationship of age and sex to stability while controlling for study and level of cooperativeness score. Age and sex were not significantly related to stability. However, level of cooperativeness was strongly coorelated with stability (r=.82). There was more variability in behavior for subjects with higher cooperation scores. This reflects the relatively low base rate for cooperative behavior -- subjects with low cooperation scores tended to have many zero per occasion (day) scores; while subjects with higher scores had more opportunity for variability.

Summary and Conclusions

Stability of cooperative social behavior in the nursery-school settings studied here was relatively high as indicated by the variance component for subjects of 15% and 30%

of total variance. Occasion differences were small and there was a relatively large subject-occasion interaction (confounded with error). This suggests that individual difference measures useful for research purposes may be based upon naturalistic observations in such settings. Estimates of generalizability coefficients from the data reported indicate that generalizable or reliable measures may be obtained based upon a relatively small sample of observations. Stability of cooperative behavior was unrelated to sex and age, although it was highly related to the amount of cooperative behavior. The generalizability analysis completed is useful both as a way of analyzing contributions to performance and estimating the size of behavior samples needed for other studies of this type of social behavior.

References

Bloom, B. I. Stability and change in human characteristics. New York: Wiley, 1964.

Cairns, R. The analysis of social interaction. New Jersey: Lawrence Erlbaum Assoc., 1979.

Cairns, R. & Green, J.A. How to assess personality and social patterns: Observations or ratings? In Cairns, R. The analysis of social interaction. New Jersey: Lawrence Erlbaum Assoc., 1979, p. 215.

Cronbach, L., Lee, J., et. al. Dependability of Behavioral Measurements. New York: Wiley, 1972.

Kagan, J. The three faces of continuity in human development. In Goslin, D.A. Handbook of socialization theory and research, Chicago: Rand McNally and Co., 1969, p. 983.

Mischel, W. Continuity and change in Personality. American Psychologist, 24, 1969, 1012-1018.

Radke-Yarrow, M. & Zahn-Waxler, C. Observing interaction: A confrontation with methodology. In cairns, R. The analysis of social interaction. New Jersey: Lawrence Erlbaum Assoc., 1979.